# 机器之心技术周报

**知识工程 知识图谱**

2019年10月19日（总第3期）第3期

本期相关主题： 注意力图卷积网络，知识图谱嵌入，BERT神经网络，生产对话系统，图数据库

本周知识图谱领域发表了多篇论文，为知识图谱中应用中出现的挑战提供了发人深醒的思路和行之有效的解决方案。例如在问答系统任务上，引入基于本体的语义成分增强正则化方法，大大提高了准确率；颠覆传统费时的静态知识图谱嵌入模式，用动态知识图谱嵌入实现高效，实时地嵌入；解决知识图谱图查询由结构不一致导致的匹配错误，系统响应时间过长的问题；成功使得面向非目标的生成对话系统能根据问题提供有充分依据的答案等。在理论方面，一种基于范畴论的新形式框架也被提出，为RDF和SPARQL的主要基本特征提供了清晰简洁的形式定义。在学习资源方面，本周推荐了一份基于python的spaCy库，手把手实现维基百科知识图谱创建的讲义。而开发资源方面，也介绍了几个有效的小工具，比如集成网络和动态推理组装器，知识图谱的可视化和知识模式建模等。更详细内容，请参照本周更新。

# 本周概览

### 热点事件 | 同盾科技云图为智能反欺诈、信贷风控和智能决策提供了强有力的科技支撑。

### 论文更新 | Efficiently Embedding Dynamic Knowledge Graphs

### 论文更新 | Bridging the Knowledge Gap: Enhancing Question Answering with World and Domain Knowledge

### 论文更新 | Semantic Guided and Response Times Bounded Top-k Similarity Search over Knowledge Graphs

### 论文更新 |On foundational aspects of RDF and SPARQL

### 学习资源 | Knowledge Graph – A Powerful Data Science Technique to Mine Information from Text

### 开发资源 | AmpliGraph (knowledge graph library)

### 开发资源 | Grakn Workbase – An Integrated Environment

### 开发资源 | Integrated Network and Dynamical Reasoning Assembler

### 近期活动 | Graphorum – 知识图谱专题演讲（Oct.14，芝加哥·美国）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 类别 | | 标题 | | | 时间 | |
| SOTA | | [Attentive Knowledge Graph Embedding for Personalized Recommendation](https://arxiv.org/abs/1910.08288) | | | **2019.10.18** | |
| 知识图谱（KG）被证实是高质量推荐的有效工具。然而，现有方法对知识图谱中用户-项目匹配的多条路径研究是独立进行的，因此无法完全捕获KG的丰富语义和底层拓扑结构。因此，论文提出了一种新颖的注意力知识图谱嵌入（AKGE）框架，以利用KG的复杂子图来链接用户-项目对，以帮助更好地推断用户的偏好。具体来说，AKGE首先采用距离感知采样策略来自动提取高阶子图，这些子图包含了具有丰富语义的用户-项目对。接着，由所提出的关注图神经网络对子图进行编码，以帮助学习准确的用户偏好项。广泛的验证表明，AKGE始终优于榜单最先进技术，另外，它还为推荐结果提供了可能的解释。  *推荐系统，注意力机制，知识图谱嵌入* | | | | | | |
| SOTA | | [Question Answering over Knowledge Graphs via Structural Query Patterns](https://arxiv.org/abs/1910.09760) | | | **2019.10.22** | |
| 知识图谱上的自然语言问答是一项重要而有趣的任务，它能使普通用户可以轻松，直观地获得准确的答案。然而，非结构化问题和结构化知识图之间的差距仍然是问答任务的一大挑战。为了解决该问题，一个很直白的解决思路就是构建结构化查询来表示输入问题，从而在知识图谱上可以进行结构化查询以生成问题的答案。与基于语义解析或模板的现有方法不同，论文提出了一种以结构查询模式提供支持的方法。首先，给定输入问题，该方法生成与知识图谱的基础结构兼容的查询草图。接着，在结构化查询模式的指导下标记节点和边来完成查询图。最后，通过在知识图谱上执行构造的查询图来检索答案。对三个回答问题基准的评估结果表明，论文提出的方法是明显优于榜单最优方法的。  *知识问答，结构化查询* | | | | | | |
| 热点事件 | | [**跨句多元关系抽取**](https://mp.weixin.qq.com/s?__biz=MzI0MTI1Nzk1MA==&mid=2651678237&idx=1&sn=f6490894df2fade7b27904e71273d56b&chksm=f2f79aacc58013ba5f038b4a5648e349fcba8e0abdb531537892ec85a1e31ac3e2212d4a5a34&mpshare=1&scene=1&srcid=&sharer_sharetime=1571955911047&sharer_shareid=9c39091165127946ebdaa4de6fa7ad2d&pass_ticket=e8eTLUU%2F3HJbRmS4RnZ0766xTiKQCLHQWD8yu%2BiJa5rLcAfuo8tHQWPzNaDumZF9#rd) | | | **2019.10.18** | |
| 关系抽取是从自由文本中获取实体间所具有的语义关系。这种语义关系常以三元组 <E1,R,E2> 的形式表达，其中，E1 和E2 表示实体，R 表示实体间所具有的语义关系。然而，关于关系抽取一个比较大的难点是跨句多元关系不能很好解决。为解决这一问题，文章介绍了几篇提供解决思路的相关论文，包括远程监督实现跨句关系抽取，图LSTM框架，图状态LSTM框架模型，给跨句多远关系抽取提供了思路。  *关系抽取，图LSTM，图状态LSTM，远程监督* | | | | | | |
| 热点事件 | | [**2019知识图谱论坛活动总结**](https://www.linkedin.com/pulse/1st-knowledge-graph-forum-2019-thomas-teske) | | | **2019.10.21** | |
| 2019知识图谱论坛已经圆满结束，此次论坛主题主要是关于知识图谱的实际应用。这份活动内容提炼主要表达了一个观点，大多数由数据驱动的企业都是研究如何从大量数据提取有效信息而不是从浓缩的知识开始的，简单观察是：我们倾向于解释，解决方案是如何根据第一原理创建的，但是从解决方案开始可能会更聪明。从而，解释解决方案的工作原理非常有力且有趣的，这也是能从此次活动得到的最大收益。  *数据驱动，知识图谱应用* | | | | | | |
| 热点事件 | | [**腾信技术：知识图谱应用**](https://mp.weixin.qq.com/s?src=11&timestamp=1571941142&ver=1932&signature=Ylea*g9c3w9f9pk*osFTIuBNugD-PLRubIa*NJpI9AuJNsBAucUsoOJecunBWHowqE36r-8BntSjfDlZez1IPfLfPYaAInYe7BlKVoOFi2AK87YsD1SpS93cS0-tItxJ&new=1) | | | **2019.10.24** | |
| 腾信自主研发的智能招采平台主要根据全国政府机关及事业单位门户网站、政府集采网站、行业招投标平台及其他媒体发布的招标中标相关信息，生成知识图谱。用户在使用平台时可根据地域、应用场景、行为习惯等自定义个人工作面板，预置的机器学习模型演练并迭代招采模型。因而用户通过智能推送此类信息的方式，能大大减少检索、分析的人工成本，提高了工作的效率及精准度。同时，智能招采平台还能通过关系提取，NLP自然语言解析处理招标、中标信息内容，以制作关系图谱，为市场行为提供知识连接新渠道与方向。  *招采知识图谱，关系图谱* | | | | | | |
| 热点事件 | | [KMS Lighthouse Extends Its Knowledge Base Integration to Freshworks Marketplace](https://www.prnewswire.com/news-releases/kms-lighthouse-extends-its-knowledge-base-integration-to-freshworks-marketplace-300944811.html) | | | **2019.10.24** | |
| 论文介绍了一种基于本体的语义成分增强正则化（OSCAR）方法。具体是在神经网络预训练过程中，嵌入来自本体或知识图谱，但与问答系统任务无关的知识。论文的数据集选取了维基百科文章，模型采用了BERT的预训练模型，任务为两个问答系统，其中一个以通用知识为主，另一个以医学健康领域专业知识为主。在BERT微调过程中，论文在引入了OSCAR后，通用知识问答系统取得了33.3%的准确率，而医学健康知识问答系统取得了18.6%准确率。总体而言，准确率比不引入OSCAR提高了4%，创造了新的榜单记录。  *知识管理，知识库* | | | | | | |
| 论文更新 | | [Using Local Knowledge Graph Construction to Scale Seq2Seq Models to Multi-Document Inputs](https://arxiv.org/abs/1910.08435) | | | **2019.10.18** | |
| 基于查询的开放域NLP任务需要从大量并且多样的Web结果中进行信息合成。为应对这类数据，当前的方法选择使用诸如TF-IDF先对数据进行排序，从中提取Web文本的一部分作为序列对序列模型的输入。区别于此，这篇论文提出为每个查询构建一个本地图结构的知识库，用以压缩Web搜索信息并减少冗余。论文表明，通过将图线性化为结构化输入序列，模型可以在标准的“序列对序列”设置内对图形表示进行编码。而实验结果表明，对于需要长文本输入，长格式问题解答和多文档摘要的生成任务，图形表示作为输入要比使用检索到的部分文本作为输入更有效。  *开放域NLP，序列对序列* | | | | | | |
| 论文更新 | | [Towards Learning Cross-Modal Perception-Trace Models](https://arxiv.org/abs/1910.08549) | | | **2019.10.18** | |
| 表征学习是最前沿深度学习方法的关键要素。它可以将原始数据转换为结构化向量空间嵌入，此类嵌入能够捕获其上下文的分布语义，例如通过自然语言句子上的单词窗口，知识图谱上的图或图像上的卷积进行嵌入。目前，此类上下文是通过人工定义的，从而产生了针对特定任务（如链接预测）的计算性能而专门优化的启发式方法。但是，这种上下文启发式模型与人类捕获信息的方式从根本上是不同的。例如，当阅读多模态网页时，1.人们不能同时感知文档的所有部分：某些单词和图像的某些部分被跳过，而另一些单词被重新访问了几次，从而感知是不连续的。2.人会在布局和设计元素的引导下在文本和图像之间转移注意力，在这些因素的影响下理解文档意思。论文实证研究了人类感知与基本嵌入模型的上下文启发式方法之间的差异。通过眼动追踪实验，捕获人类感知包含文本和图像的媒体文档的基本特征。在此基础上，论文设计了一个原型计算感知跟踪模型，称为CMPM，并评估了CMPM是如何改进基本图跳转嵌入的方法。结果表明，人类启发式计算感知模型有很大的潜力可以改进嵌入效果，这得益于模型能捕获多种形式以及布局和设计元素的特性。  *表征学习，深度学习，上下文启发模型* | | | | | | |
| 论文更新 | | [**长文综述：中文知识图谱构建技术及其应用**](https://mp.weixin.qq.com/s/DnxyrIcLqgNH5aAN1grgoA) | | | **2019.10.20** | |
| 随着智能技术的不断发展，作为人工智能支柱的知识图谱以其强大的知识表示和推理能力受到了学术界和产业界的广泛关注。近年来，知识图谱在语义搜索、问答、知识管理等领域得到了广泛的应用。构建中文知识图谱的技术也在迅速发展，不同的中文知识图谱以支持不同的应用。同时，我国在知识图谱开发方面积累的经验对非英语知识图谱的开发也有很好的借鉴意义。论文以典型的中文知识图谱为例，介绍了中文知识图谱的构建技术及其应用，此外，还对构建中文知识图谱的技术细节进行了研究。  *中文知识图谱，语义搜索，知识管理* | | | | | | |
| 论文更新 | | [Building Dynamic Knowledge Graphs from Text-based Games](https://arxiv.org/abs/1910.09532) | | | **2019.10.21** | |
| 论文的关注点是如何用文本更新知识图谱（KG）。在初步工作中，论文提出了一种新颖的序列对序列（Seq2Seq）架构来生成基本的KG操作。同时，论文还引入了一个新的基于文本的游戏状态转移（超过30万个数据点）数据集，以此来提取KG。 在完成KG提取后，该项目的下一步会尝试实现在玩基于文本的游戏时利用KG对模块进行更新。  *序列对序列，文本更新知识图谱* | | | | | | |
| 论文更新 | | [**使用预训练深度模型和迁移学习方法的端到端模糊实体匹配**](https://mp.weixin.qq.com/s?__biz=MzU2NjAxNDYwMg==&mid=2247486712&idx=1&sn=3f0f89b719e2dc17a049ea58ba8cb581&chksm=fcb3ab1dcbc4220b22aef260c83a77cd5be1a612ae11e1c6243e55ecfb225840adf2dcf10d1b&mpshare=1&scene=1&srcid=&sharer_sharetime=1571750448562&sharer_shareid=9c39091165127946ebdaa4de6fa7ad2d&key=d6525bec48886becd6e3ad3612b6ae297e757b7150754085e796471131c9e4e4c24123239cc138bbf27d6df3c41e394f231f3a2d68fa69d1a880eba3e340701b6bd00f419ed8d22921d17ecd079f478e&ascene=1&uin=MTI0NDE1NTc2MQ%3D%3D&devicetype=Windows+10&version=62060833&lang=en&pass_ticket=3%2FoGh7%2B7SFlkpPjHWK419OjhknK614%2Fw1K2rhugwLntoAefXcnZQLKfvcvaPPn9Z) | | | **2019.10.21** | |
| 目前实体匹配过程中实体之间的差异比较微妙，不同的情况下可能会有不同的决策结果，导致难以做出精确的匹配决策。另外现存的实体匹配方法在做出决策之前，往往需要大量的训练数据，而这在许多的应用场景中是难以做到的。而论文设计了一种层次化的深度模型，利用了字符级别和单词级别的信息，来预训练常见属性类型的相关模型。此外，还使用了迁移学习的方法，能够利用预训练模型并进行微调，使其能够处理新的属性类型的实体匹配任务。  *实体匹配，层次化的深度模型* | | | | | | |
| 论文更新 | | [Towards Combinational Relation Linking over Knowledge Graphs](https://arxiv.org/abs/1910.09879) | | | **2019.10.22** | |
| 给定一个自然语言短语，关系链接旨在从基础知识图谱中找到一个关系（谓词或属性）以匹配该短语。它在许多应用程序中非常有用，例如自然语言问答，个性化推荐和文本摘要。可是，传统的关系链接算法通常只为输入短语生成单个关系，而很少关注更一般和更具挑战性的问题，如组合关系链接提取了子图模式以匹配复合短语（例如岳母）。为解决这个问题，论文更专注于知识图谱上的组合关系链接的任务，由此设计了一种基于数据驱动关系组装技术的系统方法，并使该方法在元模式的指导下进行。此外，还将引入外部知识以增强系统的理解能力。最后，大量实验结果也证实了这个方法的有效性。  *组合关系链接* | | | | | | |
| 论文更新 | | [Knowledge Map: Toward a New Approach Supporting the Knowledge Management in Distributed Data Mining](https://arxiv.org/abs/1910.10547) | | | **2019.10.23** | |
| 分布式数据挖掘(DDM)指在具有分布式数据和计算的环境中寻找知识模式或模型。如今，不同地理分布和组织拥有的大量文本数据被充分挖掘，因此，产生了大量的知识。而地域的限制会导致在分析数据时，需要最小化通信。对于这个问题，现有的DDM技术分别对个别站点的本地数据进行分析，然后通过聚合这些本地结果生成一个全局模型。这两个步骤并不是相互独立的，因为原生的局部分析方法可能会产生不正确和模糊的全局数据模型。因而，这两个步骤的集成与协作需要一个有效的知识管理，也即是有效的知识图谱，以便利用挖掘出的知识对数据建模进行指导。论文由此提出了“知识地图”，一个关于挖掘知识的知识表示。该方法旨在对网格等大规模分布式平台上的知识进行有效管理。该知识图不仅方便了挖掘结果的可视化、评价，而且可以对局部挖掘过程和已有知识进行协调，提高最终模型的准确性。  *分布式数据挖掘，知识地图* | | | | | | |
| 学习资源 | | [**超全整理知识图谱相关学习资料，提供系统化的知识图谱学习路径**](https://mp.weixin.qq.com/s?src=11&timestamp=1571942284&ver=1932&signature=klevfoJvNsQh1rosBWD2kP4K-0TQpRpiogvvreyjjFYfyUqM-XcHi5-CPSN97q859y7GlgeVIPACehXv3GGEpq9mzW58P0bIlnFyDG*4wbb*egJ1gZYyKw7Aq6ISbh3X&new=1) | | | **2019.10.18** | |
| 这是一份系统化整理知识图谱相关资料的文档，为学者提供了一条学习知识图谱的路径。内容包括理论及论文，图谱及数据集，工具及服务，白皮书及报告，机构及人物，视频课程，专栏合集，评测竞赛，项目案例和相关的推广技术文章。  *领域知识图谱* | | | | | | |
| 学习资源 | | [**知识图谱的关键技术及其智能应用**](https://mp.weixin.qq.com/s?src=11&timestamp=1571942284&ver=1932&signature=S3CCM-b5bmyM5iwrz2QgkKqk0ZAJBU59jN4L-B6RSE5gDPDe4*470EKAoO60YQXY7zqc4jb-j0hp3WfB*2NuXq1B-JVMWRH7DTOzX7Gro-dDbx6apT7FJJdtH9WNDEO0&new=1) | | | **2019.10.22** | |
| 这份PPT资料来自北京大学的赵东岩老师在计算所做的《知识图谱的关键技术及其智能应用》的讲座。内容主要包括：1.从知识图谱构建、补全及其人机交互问答等方面系统阐述知识图谱的关键技术方案；2.结合北大计算机研究所在这个研究方向上的具体进展；3.以实际应用为背景、探讨如何基于知识图谱实现智能问答等智能应用的技术路线。  *人机交互问答，知识图谱构建和补全* | | | | | | |
| 学习资源 | | [**达观数据：知识图谱与语义分析技术介绍**](https://mp.weixin.qq.com/s?src=11&timestamp=1572034614&ver=1934&signature=wokfWQQULm4peEDQjzbq4iBBkW0oi6FhA-WpDaT*Zh1PEuls3Icf*ILOzHYqbll6MoVqvnzHUJi8bvwHvkCoyD9l5H6RbPiIY7GLkzkyv7-0YQpHUMu8yzgS3FrtEJ0P&new=1) | | | **2019.10.25** | |
| 本资料对知识图谱进行了一个全面的叙述，包括知识图谱的介绍，发展，主要应用，并对部分筛选的前沿论文进行了解读，使读者了解学术界和工业界关于知识图谱在QA问答中的一些新的方法。  *问答系统，语义分析* | | | | | | |
| 开发资源 | | [Pykg2vec: Python Library for KG Embedding Methods](https://github.com/Sujit-O/pykg2vec) | | | **2019.10.20** | |
| Pykg2vec是一个正在持续开发中，用于学习知识图谱中实体和关系的表示的库。这个库最大的亮点是整合了几乎将所有最新的知识图谱嵌入算法和知识图嵌入任务流水线中的必要构造块，使得这个库除了实用价值外，还有很高的学习价值。  *关系表示，实体表示，知识图谱嵌入* | | | | | | |
| 开发资源 | | [THUEE-CMCC Dialog System V1.0](https://github.com/594zyc/CMCC-KG-DialogSystem) | |  | **2019.10.24** | |
| 基于知识图谱进行套餐推荐、套餐信息查询、个人信息查询等任务的中移动智能对话系统。系统使用了三个识别器：用户动作识别器、属性识别器和值识别器，模型主要使用了 词袋模型(BOW)和富对话状态跟踪模型 (EDST)。  *知识工程，知识模式，知识图谱可视化* | | | | | | |
| 近期活动 | | [**国际顶级学术会议CIKM2019北京开幕在即，重量级嘉宾带你窥探人工智能前沿**](https://www.jiqizhixin.com/articles/2019-10-23-12) | | | **2019.11.03-11.07** | |
| 第28届国际顶级“信息提取，知识管理和数据库” 年会（CIKM - The ACM International Conference on Information and Knowledge Management）将于今年11月3日-7日在中国北京召开。自1992年成立以来，CIKM致力于将这三个相关领域的研究员和开发者聚集在一起，实现跨领域的深入合作与交流。CIKM今年以“未来生活的人工智能”为主题，聚焦大数据和人工智能未来愿景的技术和见解。此次落户北京，CIKM为国内研究和工程人员提供了一个绝佳的机会和平台去参与发表、介绍和讨论高质量的理论和应用研究成果，发现行业所面临的重要挑战，提出高效的解决方案，并共塑未来。  *知识管理，数据科学，信息检索* | | | | | | |